



A kutatási módszer és mintavétel megválasztása a tudományos kutatásokban

Dr. BELLA TAMÁS

jogász

E-mail: samatalleb@gmail.com

DOI 10.23716/TT0.22.2018.18

Absztrakt:

Bevezetés: A tudományos kutatások eredményeiből készült statisztikák torzítása, pontatlansága számtalan okra vezethető vissza, ezek közül a leggyakoribbak a mintavételi eljárások és a módszerek megválasztásakor fellépő hibák.

Célkitűzés: A mintavételi eljárások és módszerek vizsgálata azért fontos, hogy a feltárt hibaforrások kiküszöbölésével minél pontosabb statisztikai adatokhoz juthassunk.

Módszer: A mintavételi technikák és módszerek elemzése, az esetleges hibaforrások példákkal való illusztrálása a jobb megérthetőség kedvéért.

Eredmények: A valószínűségi mintavétel a legjobb módja a vizsgálandó elemek kiválasztásának, lehetőséget ad a *mintavételi hiba* becslésére, és reprezentatív arra az alapsokaságra, amelyből vették. Mintavételi hiba csökkenthető a *mintanagysággal* és a vizsgálati populáció *homogenitásának* növelésével.

Megbeszélés: Vizsgálódásaink megkezdése előtt csak olyan kérdéseket tegyünk fel, amikre tudunk, és amikre akarunk is választ adni, és csak annyi választ adjunk, ami a feltett kérdést lefedi. Ehhez a helyes mintavétel és a megfelelő módszer kiválasztása sokat segíthet.

Következtetések: Nem a mintavételi technikák, illetve a vizsgálati módszerek önmagában való alkalmazása, hanem a nem megfelelő alkalmazás torzíthatja az eredményeinket.

Kulcsszavak: mintavételi technika, vizsgálati módszer, hiba

Mottó: „Csak annak a statisztikának hiszek, amit én hamisítottam”¹

Bevezetés

A fenti mottót a statisztikában, mint tudományban kételkedők gyakran felhasználják munícióként, ha valamilyen, a statisztika módszerével készült eredményt kell kétségbe vonni. Egy tudományos kutatás eredményeiből készült statisztika esetleges

¹ Sir Winston Churchill

torzítása, pontatlansága számtalan okra, hibára vezethető vissza. Ezen okok, hibák között gyakran szerepel a mintavételi eljárás és a kutatási módszerek nem megfelelő megválasztása.

Hihetünk – e a statisztikáknak? Összefügg – e a hibás statisztikák készítése a mintavételi technikákkal és a vizsgálati módszerekkel? Ahhoz, hogy megértsük a mintavétel és a módszer jelentőségét a kutatásokban először meg kell határozni a **statisztika fogalmát**, tartalmát. A közlemény ezt követően a szakstatisztika és a statisztika, mint szakmai segédtevékenység közötti összefüggésekre világít rá azért, hogy még mélyebb részletességgel tudja feltárni az egyes mintavételi technikák és a valószínűségi mintavételhez kapcsolódó teljes felvételi hibák lehetőségeit. A könnyebb megértés kedvéért az egyes hibaforrásokat az életből, azaz a különböző tudomány területen végzett kutatások alapján vett példákkal támasztja alá. A kutatások másik leggyakoribb hibaforrása a vizsgálati módszerek helytelen alkalmazásának a területe, amire hasonlóan az életből vett példákkal világít rá a jelen közlemény.

A tanulmány akkor tölti be a célját, ha sikerül felhívnia a figyelmet arra, hogy csak a megfelelő mintavétellel, módszerekkel kaphat olyan torzításmentes eredményt a kutató, amely bizonyára meg fogja őt győzni a statisztika hasznosságáról, fontosságáról.

A statisztika kialakulása és fogalma

Eredetileg a statisztika matematikai eszközöket igénybevevő államháztartást jelentett, vagyis azon módszerek gyűjteményét, elméletét, amelyeknek segítségével a mai modern államok számon tarthatták az erőforrásaikat és a társadalmi problémákat. Ezt igazolja a szó etimológiája: A *statisticum collegicum* („államtanács”) és az olasz *statista* („államférfi, politikus”) kifejezésekből származik. A statisztika atyja Gottfried Achenwall is ebben az értelemben használta még, a mai jelentését („adatgyűjtés és adatfeldolgozás általános tudománya”) a XIX. sz. elején nyerte el. [1]

A statisztika fogalmára többféle definíció (180) létezik, amelyek közül jelenleg is különböző megfogalmazásokat használnak. Freedman szerint a statisztika annak a mestersége, hogyan lehet bizonyos rejtelmes kérdésekkel kapcsolatosan számszerű következtetésekre – vagy inkább sejtésekre – jutni. [2] Más: A statisztika olyan tudományos módszertan, illetve gyakorlati tevékenység, ami arra szolgál, hogy a valóság tényeinek valamely adott körét tömören, a számok nyelvén jellemezze. Az információ kinyerése, mint érték jelenik meg, de azzal is számolnunk kell, hogy számos információt el is veszünk a statisztikával. Babbie könyvében az alábbi megfogalmazás található: „A statisztika pedig az alkalmazott matematikának az ága, amely különféle kutatási elemzések elvégzésére alkalmas.” [3] S még

folytathatnánk a sort, de vegyük észre, hogy a különböző meghatározások lényegében két ismérven megegyeznek: A statisztika egyfelől *gyakorlati tevékenység*, másfelől *tudomány*. [4]

Az, hogy gyakorlati tevékenység, különösebb magyarázatot nem igényel, hiszen mind a mai napig használjuk, alkalmazzuk a statisztikát. (ábra 1.) A kérdés inkább az lehet, hogy tudomány e? A statisztika *tudománnyá fejlődésének folyamatát* tükrözik az alábbi, időrendben bekövetkező álláspontok:

Nem, mert nem tár fel statisztikai törvényszerűségeket.

Nem, de tudományos módszer. – Ezen elméletek művelői szerint a tudomány a természetről, a társadalomról felhalmozott bizonyítható ismeretek összessége. Az pedig, hogy miként lehet az ismereteket összegyűjteni, összefoglalni, rendszerezni önmagában nem tudomány, tehát a statisztika sem lehet az.

Igen, társadalomtudomány. – Tárgya a társadalmi jelenségek mennyiségi mozzanatainak vizsgálata.

Igen, önálló módszertudomány. – Nem egyszerűen csak a módszerek gyűjteménye, hanem saját alapfogalmakkal, definíciókkal, önálló szaknyelvezettel rendelkező egységes tudományág. [4]

A statisztika területei

A statisztika tudomány középpontjában az *általános statisztikai ismeretek* állnak. A *szakstatisztikák* az általános statisztika alkalmazásai, és mint rész tudományterületek a statisztika tudományon belül találhatók. Másrészt a statisztikai gyakorlat túlnyomóan a különböző szakstatisztikák művelését jelenti. (3,4) A statisztika különböző ágait, a szakstatisztikákat tehát nem csak például a matematikával rokoníthatjuk, hanem ugyanolyan alapon más tudományágakkal is, így például a népesedéssziszatiztikát a demográfiával, az igazságügyi sziszatiztikát a jogtudománnyal, az egészsésgügyi sziszatiztikát az orvostudománnyal, agrárstiszatiztikát az agrártudománnyal. Szakstiszatiztikaként az egyik feladat lehet az eredményeket egy irányban torzító hibák azonosítása és orvoslása az elmélet módszereivel.

A statisztika tudomány alkalmazása részben főtevékenységként, részben segédtevékenységként jelenik meg a mindennapi gyakorlatban. Az előbbieken a fő cél magának a statisztikai adatoknak létrehozása. Az utóbbiban a statisztika egy adott tudományon belül a kutatási célok elérésére, az eredmények értékelésére szolgál. (1. ábra)

Mintavétel

A statisztikai mintavétel a statisztikai gyakorlatnak az a része, amely során az alap sokaságból elemeket választunk ki, valamilyen technika mentén, azzal a szándékkal, hogy ismeretekhez jussunk a megfigyelni kívánt alap sokaságról és ennek alapján

statisztikai következtetéseket tehesünk. A mintavétel a statisztikai adatgyűjtés fontos aspektusa. A kutatók ritkán végeznek felmérést a rendelkezésre álló teljes alapsokaságon, még pedig egyrészt azért, mert nagyon magasak lennének a költségek, másrészt meg azért, mert a vizsgált sokaság elemszáma nagyon nagy, így a hozzáférés szinte lehetetlen lenne. (Példa 1, 2; 2. ábra)

Példa 1.: Dél – Afrikában egy geológus vizsgálni szeretné az év egy adott időszakában bányászott 5-10 karátos gyémántok olyan tulajdonságát, amely az adott gyémántok roncsolása nélkül nem lehetséges. Sem anyagilag, sem technikailag nem jöhet szóba, hogy az adott időszakban bányászott a fenti kritériumoknak megfelelő összes gyémántot megvizsgálják. Ekkor jöhet szóba a mintavétel.

Példa 2.: Egy időintervallumban egy adott területen (pl.: megye, közigazgatási terület, házi orvosi körzet) a lakóság szűrése, amit lehet korcsoportok, vagy nem szerint is osztályozni.

A mintavételnek két alaptípusa ismert, a valószínűségi és a nem valószínűségi mintavétel, amelyekben belül további fajtaikat ismerünk. (2. ábra)

A.) Valószínűségi mintavétel

Fő ismerv: Az alapsokaságot alkotó minden egyes elemnek azonos esélye van bekerülni a mintába. Többek között azért, mert létezik vagy készíthető az elemekből egy lista (*mintavételi keret*), ahonnan különböző technikai eljárásokkal választhatunk. (4,5)

a) Egyszerű, véletlen mintavétel

Egy adott listáról minden elem kiválasztásánál véletlen szám generátor / véletlen szám táblázat alapján választunk. (Példa 3.)

Példa 3.: A betegnyilvántartásban (központi, adott kórház, vagy csak egy bizonyos osztály) adott időpontban rögzített betegek közül véletlen szám táblázat segítségével történő kiválasztás (pl.: 1., 15, 17, 25, 78, 237) egy adott vizsgálat (pl. depresszió felmérése) elvégzése céljából. Fontos azonban azt tudni, hogy ezek az eredmények az ország, az adott kórház, vagy osztály területi kötelezettségére lehetnek reprezentatívak. Amennyiben feltüntetjük az adatok közlésekor, hogy a mintát milyen területi korláttal vettük fel, akkor ez arra reprezentatív lehet. A jelen mintavételkor azonban további hibaforrások fordulhatnak elő: A nyilvántartás nem tartalmazza azokat a betegeket, akik nem az állami, hanem a magán szektorban jelentkeztek a panaszukkal, illetve azokat sem, akik sehol nem jelentkeztek kezelésre, de betegek.

b) Szisztematikus mintavétel

Egy adott listáról csak az első szám kiválasztásánál használunk véletlen szám generátort, a továbbiakban pedig a lista minden *k*. elemét választjuk be (szisztematikusan) a mintába (*véletlen kezdőpontú szisztematikus mintavétel*). (Példa 4.)

Példa 4.: A naplóban szereplő diákok névsorából a tanár az aznapi dátum szerint választ kezdőpontot, tehát ötödikén az 5. 15. és 25. helyen szereplő diák felel.

A szisztematikus mintavételben az egyik leggyakrabban előforduló hibaként említendő az, ha az elemek bizonyos fajta elrendeződésben (*periódikusan*) szerepelnek. Ugyanis ha az elemek felsorolása ciklikus és a ciklus hossza egybeesik a mintavételi intervallum hosszával, akkor a minta erősen torzíthat.[3–5] (Példa 5.)

Példa 5.: Fiúk és lányok állnak párban. Ha ebből a mintából minden negyediket vizsgálnánk, akkor azt hinnénk, hogy csak lányokból áll a minta. Kiküszöbölhető a hiba, ha a páros és a páratlan csoportból is mintát veszünk.

c) Rétegzett mintavétel

A vizsgálni kívánt populációt több *homogén* részcsoportra osztjuk és ezekből a részcsoportokból veszünk részarányosan mintát. A rétegzett mintavétellel az a cél, hogy csökkentsek a mintavételi hibát, azaz nagyobb fokú reprezentativitást érjünk el.

A mintavételi hiba két módon csökkenthető: 1.) minél nagyobb a minta nagysága, 2.) minél homogénebb az alapsokaság. Az előző példánál maradva, ha az alapsokaságban csak fiúk vannak, akkor biztosan mondhatok állítást a fiúkra nézve, viszont a lányokról semmit. Ezért ahelyett, hogy a nagy teljes populációból (egyetemi hallgatók összessége az adott egyetemen) választunk mintát, felosztjuk az alapsokaságot homogén részcsoportokra és mindegyikből arányosan mintát veszünk. [4] (Példa 6.)

Példa 6.: A Példa 5. esethez kapcsolódva, ha az egyetemisták között nem csak a nem szerint, hanem korcsoportonként is fel szeretnénk osztani az alapsokaságot, akkor felesleges az 50, 60 vagy a 10 évesek körében vizsgálni, mert nagy valószínűséggel ezekben a részcsoportokban 0 lesz a minta szám. Megjegyzendő, hogy az évfolyamok szerinti felosztás, a korcsoportos felosztást is jelenti.

d) Többlépcsős csoportos mintavétel

Míg az előzőekben rendelkezésre állt egy adott lista a mintavételhez, addig az élet több területén nem rendelkezhetünk konkrét listával. (Így például egy adott országban az összes évészavarban szenvedő betegek száma.) Az ilyen esetekben a listakészítést és kiválasztást egymás után többször kell alkalmaznunk. (5) (Példa 7., 8.)

Példa 7.: Azt szeretnénk megtudni, hogy az adott államban mekkora a diagnosztizált és nem diagnosztizált évészavarban szenvedők száma.

Példa 8.: Ugyanígy a többlépcsős csoportos mintavételt alkalmazzuk, ha arra vagyunk kíváncsiak, hogy a választások előtt egy adott város lakosságának mi a véleménye az ideális politikus személyiségéről. (6) Amennyiben 4 évvel később azt szeretnénk megtudni, hogyan alakul a lakosság véleménye ugyanebben a kérdéskörben, akkor ugyanezen mintavételi

eljárással nyert adatokkal a korábbi eredmények összehasonlíthatók, pontosan a valószínűségi mintavételnek köszönhetően.

A valószínűségi minta vételnél a *felmérési adatok torzítása* lényegében két csoportra osztható: a *tudatos – szándékos és a tudattalan – nem szándékos*. (3. ábra)

a) Tudatos – szándékos

A szándékos torzítás lehetőségeit kiváltó körülményeket vizsgálta Ioannidis munkatársaival 2005-ben. Olyan faktorokat azonosítottak, mint az akadémiai előmenetel, a kutatások finanszírozási rendszere. Tabudöngető állításokat fogalmaztak meg a kutatási eredmények megbízhatóságával kapcsolatban, amin sok kutató felháborodott. Álláspontom szerint, a Ioannidis által leírt jelenség létezik, még ha nem is általános, hanem inkább rendhagyó. Nem szabad álszentnek lenni és azt feltételezni, hogy az életnek eme területe mentes a visszasságoktól. Ugyanis mint minden az ember által használt dolog úgy a statisztika is használható jól (jó célok szolgálatába) és rosszul (visszaélésszerűen), azonban ezek az esetek Magyarországon legtöbbször a büntetőjog területére tartoznak. (Példa 9.)

Példa 9.: Ha valaki a kutatási eredményeit hamisítja meg tudatosan azért, hogy megnyerjen egy pénzzel járó pályázatot, akkor az a magyar büntetőjog szerint többek között a csalás tényállását is kimerítheti. [7]

„374. § (1) Aki jogtalan haszonszerzés végett színlelt gazdasági tevékenységet végez, és ezzel vagyoni hátrányt okoz, gazdasági csalást követ el.”

Megjegyzendő, hogy bár a jelenség létezik, de a jelen közleménynek nem tárgya, így további részletekbe a hely hiánya miatt sem bocsátkozik.

b) Tudattalan – nem szándékos

A teljes felvételi hiba két fajtája a mintavételi és a nem mintavételi hiba. (4. ábra) A „*nem mintavételi hiba*” (*szisztematikus hibák*) nem jól definiált, mert a különböző statisztikusok különböző hibafajtákat sorolnak ide. A szisztematikus hibák esetén az eredmények állandóan, lényegesen és csak egy irányban torzítanak. Az összetevőinek száma nem rögzített, mivel az újabb és újabb módszerek, technológiák és alkalmazások új vagy módosított hibaképzőket generálhatnak. Mindezek miatt a hibafajták száma pontosan nem meghatározható. (8,9) A jelen tanulmány a Martos – féle felosztást követi. (4. ábra) [8]

Dolgok természetéből eredő:

Példa 10.: A társas élőlényeknél a társas hatás (serkentés/ gátlás) befolyásolja a viselkedést, a konkrét ingerre adott választ. Egy bevezetésre kerülő kutyatáp fogyasztását akarjuk felmérni. Nem mindegy, hogy egy kutyafalkában, vagy az egymástól elkülönítve nevelkedő kutyák között tesszük ezt. A falkában nevelkedők valószínűleg többet fognak enni az adott tápból egymás versengő jelenlétére számítva, mint a külön nevelkedő egyedek.

Használt fogalmak:

Azt a folyamatot, amelynek kapcsán meghatározzuk, hogy az egyes kifejezéseken mit fogunk érteni, *konceptualizációnak* nevezzük. Tehát a konceptualizációval előállhat a konkrét, elfogadott fogalom a kutatásunk céljaira. [4,5,10] (Példa 11.)

Példa 11.: Az evészavar előfordulási gyakoriságát (prevalenciáját) szeretnénk vizsgálni. Ahhoz, hogy egyértelmű és más hasonló jellegű vizsgálatokkal összevethető eredményekhez jussunk, meg kell először határozni az evészavar kórkép mibenlétét. Nem feltétlenül a Betegségek Nemzetközi Osztályozása (BNO) által felsoroltak játszanak itt szerepet, hanem egy az adott vizsgálattal kapcsolatos, előzetesen felállított kritérium rendszer. Így feltételül szabható a bulimia vizsgálatánál a falásroham (binge) gyakorisága, a purgáció, de akár a testsúly, vagy az ebből számított testtömeg –index is. A még pontosabb eredmények eléréséhez a fogalom mérésére használandó mutatók konkrét leírása is beletartozik. Tehát azt is pontosítani kell, hogy például ki mit ért a falásroham alatt. Például egy zsemle két perc alatti elfogyasztása még nem jelenti feltétlenül ezt, de a fél óra alatt 3000 kalória már igen. A határok től – ig is megadhatók.

Az *operacionalizálás* azon konkrét eljárások kialakítása, amelyek eredményeképpen ezeket a fogalmakat megjelenítő empirikus megfigyelésekhez juthatunk. [3–5] (Példa 12.)

Példa 12.: Az előző példa folytatásaként említendő az a vizsgálat, ahol testtömeg index és az egyes diagnosztikus tünetek súlyossága, gyakorisága alapján klinikai és szubklinikai formákat különíthetünk el, ami az adott evészavar prognózisa, relapsusa és mortalitása szempontjából lehet jelentős.

Megfigyelés, mérés:

A klinikai vizsgálatoknál, különösen az eset – kontroll jellegűeknél az egyik sarkalatos pont, hogy a mintavétel torzításmentes legyen. [3,4] (Példa 13.)

Példa 13.: Közismert tény, hogy a 25 fok feletti laborhőmérsékleten levő műszerek 50%-kal magasabb értékeket mérnek. A mérési eredmények konzekvensen ennyivel magasabb értéket jeleznek, tehát az eltérés mindig egy irányba mutat, vagyis nem a műszer pontatlanságáról van szó.

B.) Nem valószínűségi mintavétel

Fő ismerv: Bizonyos kutatási helyzetekben nem létezik, vagy nem tudunk összeállítani egy listát a vizsgálni kívánt alapsokaságról, mert az nem lenne reprezentatív, ezért nem általánosítható. Viszont az így kapott eredmények érvényessége magas.

a) Egyszerűen elérhető alanyok

A kutató a könnyen elérhető elemekre hagyatkozik. Ez a módszer azért népszerű, mert egyszerű, gyors és olcsó. (pl.: egy bevásárló központ bejáratánál az utca embere, egyetemisták egy adott teszt standardizálására, vagy online felmérések)

Példa 14.: Egyszerűen elérhető alanyokra, legtöbbször az egyetemistákra támaszkodnak a kutatók egy kérdőív megszerkesztésénél, validálásánál. (Resch, M., Bella, T.: PTKS –76 = Pihentőség - Terheltség Kérdőív Sportolóknak, MCT = Melléknév Címjegyzék Teszt)

b) Szakértői kiválasztás

Bizonyos esetekben akkor kaphatjuk a kutatás kérdéseinek a megválaszolásához a legmegfelelőbb mintát, ha annak alapján választjuk ki, amit a populáció elemeiről már előzetesen tudunk.

c) Hólabda technika

A hólabda módszert akkor alkalmazzuk, ha nehézségekbe ütközik a vizsgálni kívánt populációnak már a körülhatárolása is. Egy nehezen elérhető társadalmi csoportot (*rejtett populációt*) szeretnénk vizsgálni, amelynek tagjairól nem létezik, vagy nem áll rendelkezésünkre lista, nyilvántartás. Amennyiben a vizsgálni kívánt csoport egyik, vagy néhány tagját sikerül rávenni a kutatásban való részvételre, adatokat gyűjtünk, majd megkérjük, hogy adja meg más, a célzott populációba tartozó személyek elérhetőségét. [11,12] (Példa 15.)

Példa 15.: Stigmatizált kisebbségi csoportok, hajléktalanok, droghasználók, illegális bevándorlók, az egyes vallási közösséghez tartozók, vagy a nemi identitás alapján elkülönülők. A hólabda módszerrel lehet felmérni a homoszexuálisok körében a depresszió előfordulási gyakoriságát.

Megjegyzendő, hogy az előző példában szereplő szexuális irányultságról jogszabályi tilalom alapján sem létezhet lista. (13,14) Az alapvető jogok biztosának a homofóbia és transzfóbia ellenes világnap alkalmából kiadott közleménye is kimondja ezt a tilalmat: „Az Alkotmánybíróság gyakorlata értelmében az azonos neműek tartós párkapcsolatának védelme az emberi méltósághoz való jogból, az önrendelkezési jogból illetve a személyiség szabad kibontakoztatásához való jogból következik.” (13,14)

d) Kvótás mintavétel

Kvótás mintavétel esetén reprezentatív mintára törekszünk, de sajnos nem áll rendelkezésünkre lista az alapsokaságról. Ismerjük azonban az alapsokaság struktúráját (pl. megbízható, hiteles népszámlálási adatok alapján). Kiindulópontunk egy, a sokaságot a kutatás szempontjából fontos változók szerint (pl. nem, életkor, iskolai végzettség) leíró táblázat (*mátrix*). Ha ismerjük a mátrix egyes celláiba tartozók arányát, a mintánkat úgy választjuk ki, hogy az egyes cellába az alapsokaság arányainak megfelelő számú válaszadó kerüljön úgy, hogy a válaszadók mintába kerülése véletlenszerűen történjen. Tehát a mintánk követi a teljes populációt jellemző arányokat. [3]

Példa 16.: Kvótás mintavétel segítségével jósolta meg sikeresen 1936-ban az akkor fiatal kutató, Georg Gallup és intézete Roosevelttel győzelmét az amerikai elnökválasztáson. (3)

e) *Önkiválasztás*

Az önkiválasztós minta esetén nem a kutató választja a mintát, hanem a válaszadók önként jelentkeznek részvételre. (Példa 17.)

Példa 17.: A televízióban a különböző show műsorok, vagy táncdal fesztiválok betelefonálásos, vagy SMS küldéses szavazásai tartozhatnak ide, de ugyanígy a választott média fórumtól függetlenül az internetes szavazásokra is gondolhatunk egy önkiválasztós minta esetén.

Az ilyen mintavételen alapuló kutatások úgynevezett álfelmérések, tudományos értékük nincs, mivel az így kapott eredményeket semmiképpen sem vetíthetjük ki semmilyen nagyobb populációra, akármilyen magas lenne a válaszadók száma.

Vizsgálati módszerek

A különböző vizsgálati módszereknek többféle csoportosítása létezik, amelyből az egyik lehetőség, ha két fajtára osztjuk fel: a beavatkozással járó és a beavatkozástól mentes módszerek. Mind a két csoportba tartozhatnak kvantitatív, illetve kvalitatív jellegű technikák. A lehetséges vizsgálati módszereket mutatja be az 5. ábra.

A) *beavatkozással járó*

A társadalomtudományi kutatásokban a beavatkozással járó módszereknél a kutató nem kívülről szemlélőként figyeli a történéseket, hanem valamilyen mértékben beleavatkozik a vizsgálandó folyamatba, ahol közvetlenül is kapcsolatba kerülhet az alapsokasággal. Az egészségügyben a beavatkozással járó módszereket, invazív eljárásoknak nevezik, amit a beteg írásos beleegyezésével lehet csak elvégezni, mert nem mindig jár állagsérelem nélkül.

Példa 18.: Tüdőgyógyászatban, vagy belgyógyászatban a különböző szervekből biopsziával történő szövetminta vétel későbbi vizsgálatok elvégzése céljából.

Megjegyzendő, hogy most is a korábban már részletezett konceptualizálás problematikájával állunk szemben. [3]

a) *Kísérlet*

A természettudományok területén az egyik leggyakrabban használt módszer a kísérlet. A laboratóriumi körülmények között ritka, hogy egyetlen, azaz az ingert kapó, vagy a beavatkozással járó csoportot vizsgáljuk, hanem legtöbbször van mellette egy másik, az ingertől vagy beavatkozástól mentes ún. kontrol csoport. Mindkét csoportban előzetes mérést kell végezni. Majd az egyik csoportban az inger adását követően, a másikban pedig e nélkül végezzük el a második, ún. utólagos mérést. Ezek után tudunk következtetéseket levonni a függő és a független változókra.

Példa 19.: Orvostudományban ilyenek a placebo kontrollált vak, illetve kettős vak vizsgálatok. „Vak”, amikor csak a beválasztott személy nem tudja, hogy most placebót (hatóanyag tartalom nélküli, küllemre ugyanolyan szer, mint a hatóanyag tartalommal bíró), vagy hatóanyag tartalommal bíró szert kap a vizsgálatban résztvevő. „Kettős vak”, amikor maga a beválasztó, az adatokat felvevő személy sem tudja ugyanezt, mert azt majd csak a kísérlet lebonyolításában részt nem vevő adatelemző fogja megismerni. Ez azért fontos, mert egy gyógyszer kipróbálásánál sokszor maguk az orvosok is „nagyobb javulást figyelnek meg” azoknál, akik az adott szert kapják, mint akik nem. Gyakori hiba még, ha nem végeznek előzetes, vagy utólagos méréseket mindkét (kísérleti és kontroll) csoportban, vagy nem ugyanabban a megadott időpontban, időintervallumban teszik meg.

Nagyon fontos, hogy a két csoport csak az ingeradásában, illetőleg annak hiányában különbözzön, tehát a csoportok lehető legtöbb tulajdonsága megegyező legyen.

Példa 20.: Egy fogyasztó szer vizsgálatához nem szerencsés, ha az egyik mintában valamennyi egyednek 30 és afeletti a testtömeg indexe, míg a másik csoport zöme 20-25 közötti testtömeg index értékkel rendelkezik. Tovább lehet homogenizálni a két minta csoportot az életkor, vagy a nemek aránya alapján.

Ezen hibának a kiküszöbölésére lehetne egyfajta megoldás a már korábban említett valószínűségi mintavétel, de a minta elemszám (100 alatti) alacsonyága miatt ezt nem használják, hanem helyette a gyakorlatban az un. randomizált beválasztás terjedt el. Ez azt jelenti, hogy csak egy homogén mintánk van, ahonnan az elemeket véletlenszerűen osztjuk szét a két (kísérleti, illetve kontroll) csoportba.

b) Kérdőíves vizsgálatok

A kérdőíves vizsgálatok alkalmasak leíró, magyarázó és felderítő célokra. Ezek a vizsgálatok lehetnek önkitöltősek, vagy a kérdező biztos által felvettek. A válaszadás megtagadásából, illetve az önkéntességből eredő gyakori hibákat mutatja be az alábbi példa.

Példa 21.: Ugyanúgy befolyásolja az eredményeket, ha a vizsgáltak között magas a nem válaszoló, vagy a szándékosan hibásan válaszolók aránya. Például egy kábítószer használat szövődményeit felmérő önkéntes, anonim vizsgálatban a kábítószer használat tényét, vagy a szerhasználat mennyiségét letagadó válaszok módosítják a kapott eredményeket, például azt, hogy ennek alapján bizonyos kábító szereket mennyire tartanánk veszélyesnek.

A kérdező biztosok által felvett kérdőívek a fent említett hibákat, félreértelmezéseket, válasz megtagadásokat küszöbölnék ki, mert eleve nem venne részt az összesítésben az ilyen egyed, azaz magasabb lenne a válaszadási arány. A leggyakoribb hiba, ha a kérdező biztos nem megfelelően felkészített.

Példa 22.: Egy depressziós tüneteket felmérő vizsgálatnál fontos, hogy minden kérdezőbiztosnak ugyanazt jelentsék az adott tünetek, ugyanolyan besorolást használjon.

A kvalitatív vizsgálatokhoz a terepkutatás tartozik, amelynek két fajtája létezik, az interjú és a fókusz csoport. (5. ábra)

c) Interjú

Az interjú kapcsán sokkal cizelláltabban tehetjük fel az adott kérdést, viszont nagyobb a hiba lehetőség arra, hogy nem megfelelően, nem közérthetően megfogalmazva tesszük mindezt. Néhány általános gyakorlati tanács: kerüljük az eldöntendő, vagy sugalmazó kérdéseket. Röviden, velősen fogalmazzunk és a releváns információk megszerzésére törekedjünk. A kérdező legyen kompetens az adott témában.

Példa 23.: Ilyen interjú helyzet lehet, amikor a pszichiáter explorál (első vizsgálatkor anamnézis felvétel), vagy a rendőr / bíró, amikor tanút hallgat ki egy adott eseményre vonatkozóan.

d) fókuszcsoport

A technika itt hasonló, mint a fent említett interjú során, csak most egy kiscsoport, 10–15 személy áll a kérdezővel szemben. Fontos tudatosítani, hogy nincsenek jó és rossz válaszok. Előnye, hogy flexibilis és költség – hatékony. A hátrányai (a beszélgetést elősegítő környezet megteremtése, nehéz az adatelemzés) ellenére is nagyon elterjedt napjainkban.

Példa 24.: Egy bevezetésre kerülő termékről gyűjtünk információt kis csoportokkal. Gyakori hiba, ha befolyásoljuk a válaszadást jutalmakkal, ajándékkal. Megjegyzendő, hogy ilyenkor valószínűleg az eredmények tudatosan egy irányban fognak torzítani.

B) beavatkozástól mentes

A beavatkozástól mentes módszerek kapcsán a megfigyelendő minta legtöbbször valamilyen már meglevő adat, ismeret, leírás, amiből különböző technikai eljárásokkal juthatunk további fontos összefüggésekhez, információkhoz. Három fajtájuk ismeretes. Ezekből mind kvalitatív, mind kvantitatív technikával lehet a tartalomelemzés és a történeti, összehasonlító elemzés módszereit használni. A meglevő statisztikák viszont csak kvalitatívan elemezhetők. (5. ábra)

a) Tartalomelemzés

A tudományos igényű tartalomelemzés az utóbbi két évtizedben vált ismertté hazánkban: pszichiátria, szociológia, néprajz, politológia, irodalom- és történelemtudományok használják a legszívesebben. Tartalomelemzéssel írott, halott, vagy képi formában megjelenő tartalmakat értelmezhetünk. [15,16] Ezen technika segítségével a vizsgáló azonosítja és/vagy feltárja a fókuszcsoportok, az interjúk szövegében (átírataiban), a reklámokban, vagy más szövegekben a szavak jelentéseit, a fogalmak értelmét, illetve azok összetett kapcsolatait. A szavak, szófordulatok gyakoriságából és a fogalmak egymáshoz viszonyított kapcsolataiból következtetéseket von le, hipotéziseket vet fel a közlés forrásáról. [17,18] A tartalmak lehetnek nyilvánvalóak (manifeszt), vagy mélyen lappangók (latens). Kezdetben csak a manifeszt, egyértelműen megjelenő tartalmakat vették kizárólag

MTT Tudománytörténeti Kötetek II.

figyelembe, de ma már a latens tartalmak elemzésével olyan következtetések is levonhatók, amik első látásra nem láthatók, nem észlelhetők, miáltal jóval több információhoz juthatunk.

A tartalomelemzés több lépcsőből áll:

- *Első lépés* a vizsgálandó tartalom, vagy tartalmak *kiválasztása*. Lehet egy korábbi saját kutatás keretében felvett anyag (pl.: interjúk, beszámolók, képek, felvételek) vagy mások meglevő szövegeinek (újságcikk, levelezés, napló, chat, komment), egyéb tartalmainak elemzésre történő kiválogatása egy előzetes hipotézis alapján. [18] Mind a két technikának vannak előnyei és hátrányai. A saját anyagunkban ismeretesekek vagyunk, tudjuk, hogy mit hol találunk meg, ezért a későbbi újrafelhasználás kevésbé időigényes. Míg az idegen anyag elemzése során valószínűleg ugyanezen okból nagyobb mennyiségű anyag áttekintésére lesz szükség, hiszen nem biztos, hogy a kiszemelt rész alkalmas lesz az általunk kiszemelt rész vizsgálatára, elemzésére. Lehet, hogy az adott szövegrész egy irányban elfogult, torzít, akár az adott korszak, vagy a lejegyző szubjektuma által. Tény, hogy egy nagyobb terjedelmű szövegből nehezebb a reprezentatív minta kiválasztása.
 - *Második lépés a kódolás*, amikor a különböző tartalmakat egyneműsítik, ilyen például a videó leírása, átírása. Gyakori hibaforrás, ha nem kellően részletes a leírás: a kontextus, a hangszín, a különböző nonverbális, metakommunikatív jegyek lejegyzése is fontos lehet, hiszen nem egyszer pont ellentétes értelmet adhat az elhangzott szövegnek. A kvalitatív tartalomelemzés esetén a kódolás nem előre meghatározott kategóriák szerint zajlik, mert ezek az elemzés folyamatában alakulnak ki. [16,18]
- Példa 25. Egy családterápiás ülésről készült feljegyzéshez hozzátartozik a megfigyelők által elmondottak, illetve a videó felvételen a nonverbális jegyek, metakommunikatív jelzések leírása, elemzése is. Ezek kihagyása óriási hiba, éppen azok többletjelentés – tartalma, illetve a verbális szövegek tartalmának módosítása miatt. Ha ezek a releváns információk lemaradnak, akkor az összefüggések, az új kapcsolódási pontok, következtetések és értelmezések hibásak, tévesek lesznek.
- *Harmadik lépés az elemzés*. Miután minden kód valamilyen jelentéstartalmat hordoz, a csoportosításuk valamilyen teória mentén többletinformációt hordoz. Ilyen például két-három kód rendszeres együttes előfordulása, ami az un. latens tartalomhoz tartozó többlet információval bírhat. [16, 17, 19]
 - *Negyedik lépés az értelmezés*. A tendenciaszerű együttes előfordulások olyan törvényszerűségeket mutatnak, amelyek tehát új jelentéstartalmat hordoznak, vagy éppen ellenkezőleg, ha folyamatosan bizonyos jelek, tartalmak hiányoznak, akkor bizonyos (latens) mozzanatokra lehet árulkodó.

Példa 26.: Egy személyiségrajz elkészítéséhez összegyűjtjük a már meglevő iratokat, naplókat, jegyzeteket, levelezéseket, de ugyanígy a készült fotókat, vagy egyéb hang illetve képi felvételeket. Ezeket dekódoljuk, majd a személyiségrajzot egy hipotézis mentén a jellemző karakterjegyek kiemelésével, illetve azok összevetésével, ütköztetésével készítjük el. [18]

b) Történeti, összehasonlító elemzés

A források valódiságára, megbízhatóságára vonatkozó adatokat kell számba venni, ami a *forráskritika* feladata.

Az elemzés lehet

- leíró: az eseményeket időrendben ismerteti. Az eltéréseke koncentráló „kontrasztív”, vagy az egyedi vonások, sajátosságok feltárásával „individualizáló” típusúak lehetnek.
Gyakori hiba, hogy túl nagy az összehasonlított esetek közötti eltérés, különbség, ezért az összehasonlításra lényegében alkalmatlanná válik, hiszen nem lehet tudni, hogy a változásokat konkrétan mely tényezők okozhatják.
- értelmező: az adott korszak tükrében, a teljes rendszer összefüggéseiben tárgyalja
- összehasonlító: időben (a különböző korszakokat) vagy térben (régiókat) hasonlítja össze a vizsgálandó elemeket, tartalmakat

Példa 27.: A nyelvészetben gyakran használt eljárás a történeti összehasonlító módszer, a nyelv eredetének felkutatása mellett többek között az őshazára és a korábbi történeti korok eseményeire, a népek vándorlásaira lehet belőle következtetni. [19]

- teoretikus: egyetemes jellegű értelmezésekre törekszik [15]

c) meglevő statisztikák elemzése

A meglevő statisztikák járulékos adatforrásként is hasznosak a különböző kutatásokban. A világháló a releváns elérhető adatforrások, statisztikák megtalálásában jelentős szereppel bír. A számítógépes információkereső rendszerek gyors és kiterjedt keresést tesznek lehetővé. Itt az elemzési egységek általában nem az egyének, hanem kisebb – nagyobb létszámú csoportok. Téves megállapításhoz vezethet, ha ezeket a csoportszintű összefüggéseket átkonvertálás nélkül alkalmazzuk az egyedekre is. Az érvényesség és a megbízhatóság terén pedig gyakori hiba az, ha a statisztikák nem egyértelműen, nem pontosan a kutatási kérdés megválaszolásához szükséges változókra vonatkoznak. A megalapozott logikus gondolkodás, érvelés és több független eljárás megismétlésével lehet az alapfeltevéseket megerősíteni.

Összefoglalás

S végül összefoglalásként néhány alaptézist szeretnénk megfontolásul a jelen tanulmányból kiemelni azon kutatók számára, akik komolyan értékelhető eredményekre szeretnének jutni a vizsgálataik során.

Először is fontos, hogy a vizsgálódásaink megkezdése előtt csak olyan kérdéseket tegyünk fel, amikre tudunk, és amikre akarunk is választ adni, tehát ne maradjanak nyitott, un. „költői” kérdések. Természetesen, az olyan kérdések, amik a vizsgálatok megkezdése előtt nem merültek fel bennünk, azokra ne is keressünk megoldást, csak a korábban már feltett kérdésnek megfelelő válaszokra szorítkozzunk. [18]

Néhány gyakorlati jó tanácsként kiemelhető, hogy a valószínűségi mintavétel a legjobb módja a vizsgálandó elemek kiválasztásának, mert megszünteti a kutató *részrehajlását* és lehetőséget ad a *mintavételi hiba* becslésére, másrészt pedig reprezentatív lesz arra az alapsokaságra, amelyből vették. Mintavételi hiba csökkenthető a *mintanagysággal* és a vizsgálati populáció *homogenitásával* növelésével.

A fentiek ismeretében most tegyük fel ismételten a korábbi kérdést: Hihetünk-e a statisztikáknak? – Bátran állíthatjuk, hogy igen, mert nem a mintavételi technikák, illetve a vizsgálati módszerek önmagában való alkalmazása, hanem a nem megfelelő alkalmazás torzíthatja az eredményeinket.

Irodalomjegyzék

1. BALL, P. (2004): *Critical Mass. Farrar, Straus and Giroux*. 53. ².
2. FREEDMAN, D., PISANI, R., PURVES, R. (2005): *Statisztika*. Typotex, Budapest.
3. BABBIE, E. (2003): *The Practice of Social Research*. [A társadalomtudományi kutatás gyakorlata] 6. kiadás, Balassi Kiadó, Budapest.
4. PUKLI, P., VÉGVÁRI, J. (2004): *A statisztika: tudomány és szakma*. Stat. Szle., 82, 1, 5–30.
5. PROHÁSZKA, Z., FÜST, GY., DINYA, E. (2009): *Biostatisztika a klinikumban: alapfogalmak és módszerek használata feladatok megoldására*. (szerk.: Prohászka Z.) Semmelweis Kiadó, Budapest.
6. BELLA, T. (2009): *Az (ideális) politikus személyiségének reprezentációi a különböző szocio-demográfiai változók mentén*. Szakdolgozat. ELTE, Társadalomtudományi Kar.

² ISBN 0-374-53041-6

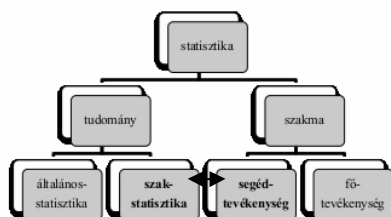
7. 2012. évi C törvény a Büntetőtörvénykönyvről
8. MARTOS Á. (2005): A mintavételi hiba kiszámítása és felhasználása a hivatalos statisztikákban. Stat. Szle., 83:613–627.
9. ROBERT, M. (2010): Groves Lars Lyberg. Total Survey Error. Past Present and Future. Public Opinion Quarterly, 74/5.
10. IOANNIDIS, J. P. A. (2005): Why most published research findings are false PLoS Medicine, 2, 8, e124.
11. SCHLEICHER, N. (2007): Kvalitatív kutatási módszerek a társadalomtudományokban. BKF jegyzet,³ Századvég, Budapest.
12. ELFIL, M., NEGIDA, A. (2017): Sampling methods in Clinical Research; an Educational Review. Emergency. 5:e52
13. Magyarország Alaptörvénye (2011.04.25.)
14. Az egyenlő bánásmódról és az esélyegyenlőség előmozdításáról szóló 2003. évi CXXV. törvény.
15. KAELEBLE, H. (1999): Comparative Historiography: Problems and Perspectives. History and Theory. 38:25-39.
16. RESCH, M., BELLA, T.: A politikusok szerepjátékai. Készítette: Méhes Ákos, Kisalföld. 2009. 08. 01. ⁴
17. PROHÁSZKA, Z.(2016): Populáció, minta és adatok: mintavételezés a klinikumban
18. *semmelweis.hu/kutlab/files/2016/01/vizsglattervezs_poweranalzis20153.ppt (előadás)*
19. RESCH, M., Bella, T. (2010): A politikus személyiségének gondolkodási és magatartási vonatkozásai. Orv. Hetil. 151:111–118.
20. KLIMA, L.: Az uráli őshaza kutatásának története.⁵

³ ISBN: 963 734053 6

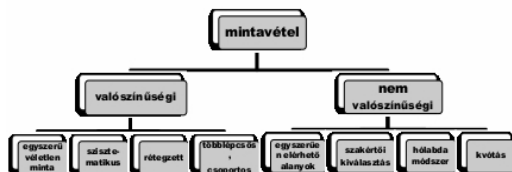
⁴ http://www.kisalfold.hu/szieszta/politikusok_szerepjatekai/2109318/ (riport)

⁵ <http://finnugor.elte.hu/?q=ostortenet> (tanulmány)

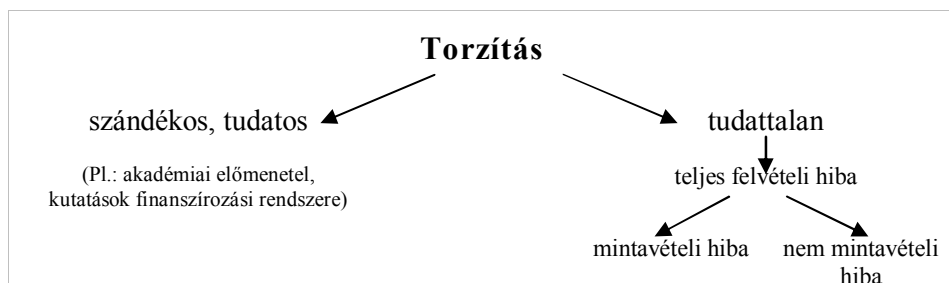
Ábrák



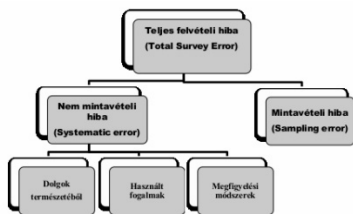
1. ábra: A statisztika területei⁶



2. ábra: A mintavételi technikák



3. ábra: Valószínűségi mintavételnél a felmérési adatok torzításának forrásai



4. ábra: A teljes felvételi hiba összetevői⁷



5. ábra: Vizsgálati módszerek⁸

⁶ PUKLIP; VÉGVÁRI J. (2004): A statisztika: tudomány és szakma. Statisztikai szemle, 82/1., nyomán BELLA T.

⁷ MARTOS, Á. (2005): A mintavételi hiba kiszámítása és felhasználása a hivatalos statisztikákban. Stat Szle, 83/7., nyomán BELLA, T.

⁸ BABBIE, E., és mtsai (2000): Adventures in Social Research Newbury Park CA, Pine Forge Press.

Selection of sampling techniques and methods in scientific research

Introduction: The bias and inaccuracy of statistics emerging from scientific research findings can be traced back to infinite causes, the most frequent of which are the errors committed upon selecting the sampling techniques and methods themselves.

Research objective: Examining sampling techniques and methods is important for us in order to obtain the most accurate statistical data possible by eliminating the identified causes of error.

Method: analysis of sampling techniques and methods and illustration of possible causes of error identified for better understanding.

Results: The probability sampling method is the best way to choose the objects to test since it allows for estimating *sampling error*, and is representative of the sample population. Sampling errors can be limited by *sample size* and by increasing the *homogeneity* of the population.

Discussion: Before starting our tests, we should ask only questions that we are able and willing to find answers to, and provide only answers that cover the question raised. This can largely be supported by identifying the right sampling method and techniques.

Conclusions: Instead of sampling techniques and testing methods themselves, inappropriate application distorts our findings.

Keywords: sampling technique, examination method, error